STATISTICAL FEATURE EXTRACTION FROM COMPRESSED VIDEO SEQUENCES

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull
Image Communications Group, Centre for Communications Research,
University of Bristol, Merchant Ventures Building, Woodland Road, Bristol BS8 1UB,
United Kingdom.
Voice - +44-117-954-5198, Fax - +44-117-954-5206
Email - W.A.C.Fernando@bristol.ac.uk

ABSTRACT

To maximise the benefits from data compression, it would be advantageous to develop algorithms that do not require decompression to extract relevant information for post processing. In this paper, a novel technique for extracting variance is proposed for MPEG-2 compressed video using Parseval's theorem. Results show that the estimated variance closely matches with the actual variance. Furthermore, this technique is applied to identify scene changes in compressed domain.

1. INTRODUCTION

A major feature required in a visual information system is an efficient indexing scheme to enable fast access to the stored data. A common and natural idea is to index the video sequences first into video shots by identifying both sudden and gradual scene changes and then to extract features.

The large channel bandwidth and memory requirements for the transmission and storage of image and video necessitate the use of compression techniques [1,2]. Hence, the visual data in future multimedia databases is expected to be stored in the compressed form. In almost all multimedia systems, there is a need to integrate compression and processing of images and video. To maximise the benefits from data compression without incurring extra computation and storage, it would be advantageous to develop processing algorithms that do not require decompression of the entire compressed data. Therefore a powerful scene change detection, which can operate both in compressed and uncompressed domain, is required to allow for a complete characterisation of the video sequences.

Statistical features have been used to identify both sudden and gradual scene changes in uncompressed video sequences [3-7]. Most of these algorithms, which are proposed for uncompressed video, can be applied for compressed video, if we can extract the relevant statistical features. Yeo [7] proposed DC-sequences to extract mean from each image in compressed video. However, variance calculation in compressed domain is difficult.

In this paper we present a novel scheme based on Parseval's theorem to extract variance from compressed video. A reduced image is formed with extracted statistical features and defined this image as the statistical image. These statistical images can then be used for higher level processing such as object tracking and scene change detection in video sequences. Rest of the paper is organised as follows: Section 2 presents how the variance is evaluated from compressed domain features. Experimental results are given in section 3. Finally, section 4 presents the conclusions and the future work.

2. ESTIMATION OF VARIANCE IN COMPRESSED DOMAIN

In general, a macroblock (MB) of size $N_1 \mathrm{x} N_2$ is considered. Let, $S(k,i,j)$ and $F(k,i,j)$ be the uncompressed image and compressed image respectively. Here, k is the MB number ($k=1:N$, N is the total number of MBs in an image) Let, $\mu_{k,s}$ and $\sigma^2_{k,s}$ be mean and the variance of the k^{th} MB in uncompressed image respectively. Then, mean and variance of the k^{th} MB can be evaluated as shown in Equation (1) and Equation (2).

$$\mu_{k,s} = \left(\frac{1}{N_1 N_2}\right)\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} S(k,i,j) \tag{1}$$

$$\sigma^2_{k,s} = \left(\frac{1}{N_1 N_2}\right)\left[\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} S(k,i,j)^2\right] - \mu^2_{k,s} \tag{2}$$

Two-dimensional DCT of a N_1 x N_2 MB can be defined as follows,

$$F(k,p,l) = \sqrt{\frac{4}{N_1 N_2}}C(p)C(l)\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} S(k,i,j).Cos\left(\tfrac{(2i+1)}{2N_1}\pi p\right)Cos\left(\tfrac{(2j+1)}{2N_2}\pi l\right) \tag{3}$$

where, $C(p),C(l) = \begin{cases}\frac{1}{\sqrt{2}} & for\ p,l=0\\ 1 & otherwise\end{cases}$

Using Parseval's theorem, power in uncompressed domain signal is related to compressed domain signal as shown in Equation (4) when the quantisation noise is negligible.

$$\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} S(k,i,j)^2 = \sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} F(k,i,j)^2 \tag{4}$$

Mean of the k^{th} MB can be calculated from Equation (3) as follows.

$$\mu_{k,s} = \sqrt{\frac{1}{N_1 N_2}} F(k,0,0) \tag{5}$$

Therefore Equations (2), (4) and (5) gives,

$$\begin{aligned}\sigma_{k,s}^2 &= \left(\frac{1}{N_1 N_2}\right)\left[\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} F(k,i,j)^2 - F(k,0,0)^2\right] \\ &= \left(\frac{1}{N_1 N_2}\right)\left[\sum_{i=1}^{N_1-1}\sum_{j=1}^{N_2-1} F(k,i,j)^2\right]\end{aligned} \tag{6}$$

Equation (6) shows that $\sigma_{k,s}^2$ can be calculated by taking the average power of AC-coefficients in each MB. Therefore, mean and variance of each MB can be evaluated in DCT compressed domain using equation (5) and (6) respectively. It can also be shown that mean (μ_s) and variance (σ_s^2) of the whole image, which contains N number of MBs, can be calculated from Equation (7) and (8) as given below.

$$\mu_s = \frac{1}{N}\sqrt{\frac{1}{N_1 N_2}}\sum_{k=1}^{N} F(k,0,0) \tag{7}$$

$$\sigma_s^2 = \left(\frac{1}{N N_1 N_2}\right)\left(\left[\sum_{k=1}^{N}\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} F(k,i,j)^2\right] - \frac{1}{N}\left[\sum_{k=1}^{N} F(k,0,0)\right]^2\right) \tag{8}$$

MPEG-2 defines three types of pictures: intra (I)-Pictures, Predictive pictures (P-Pictures) and Bi-directional pictures (B-Pictures). The extraction of statistical features from each MB in intra (I) frames is explained before. However, feature extraction for inter-coded frames (P and B) is not trivial since DCT coefficients for these MBs are not readily available. Therefore, DCT coefficients of each inter-coded MBs need to be evaluated.

2.1 DCT coefficients estimation

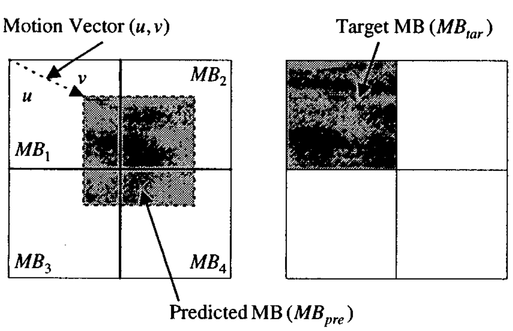

Figure 1: Graphical representation of four neighbouring MBs and motion vectors

Figure 1 shows the four neighboring MBs and motion vectors of the predicted (virtual) MB. DCT coefficients of inter-coded frames can be extracted from an intra-coded frame. The computation of the DCT coefficients of a new image block at any arbitrary position can be computed from the DCT coefficients of the four original neighboring blocks. This takes the form of pre-multiplying and post-multiplying of those blocks with appropriate matrices (S_{i1} and S_{i2}) [8]. Equation (9) describes the process of DCT coefficients evaluation for a motion-compensated MB.

$$DCT(MB_{pre}) = \sum_{i=1}^{4} DCT(S_{i1}) DCT(MB_i) DCT(S_{i2}) \tag{9}$$

There are four possible locations for the sub-block of interest (with reference to MB_i): lower-right, lower-left, upper-right and upper-left. These locations define S_{i1} and S_{i2} matrices as tabulated in Table 1. Parameters h_i and w_i are the height and width of the overlap of MB_{pre} with MB_i. For a particular MB, these two parameters can be evaluated from its motion vector (u,v). Therefore, DCT coefficients can be evaluated for the predicted MB using Equation (9). DCT coefficients of the error term (MB_{er}) are readily available from MPEG-2 compressed video. Finally, DCT coefficients of the target (current) MB (MB_{tag}) are calculated by adding the DCT coefficients of the predicted MB and the DCT coefficients of the displaced frame difference (DFD) signal as shown in Equation (10).

$$\begin{aligned}DCT(MB_{tag}) &= DCT(MB_{pre} + MB_{er}) \\ &= DCT(MB_{pre}) + DCT(MB_{er})\end{aligned} \tag{10}$$

Sub-block	Position	S_{i1}	S_{i2}
MB_1	lower right	$\begin{bmatrix} 0 & I_{h_1} \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ I_{w_1} & 0 \end{bmatrix}$
MB_2	lower left	$\begin{bmatrix} 0 & I_{h_2} \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & I_{w_2} \\ 0 & 0 \end{bmatrix}$
MB_3	upper right	$\begin{bmatrix} 0 & 0 \\ I_{h_3} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ I_{w_3} & 0 \end{bmatrix}$
MB_4	upper left	$\begin{bmatrix} 0 & 0 \\ I_{h_4} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & I_{w_4} \\ 0 & 0 \end{bmatrix}$

Table 1: S_{i1} and S_{i2} matrices (I is an identity matrix of h_i or w_i)

When DCT coefficients are extracted both mean and variance can be evaluated with Equations (7) and (8) for inter-coded frames.

Hence S-sequences can be defined for both intra and motion compensated frames from Equation (11) and (12).

$$S'_{\mu,m,n} = \frac{1}{k_1 k_2 \sqrt{N_1 N_2}} \sum_{k=1}^{k_1 k_2} F_{(m-1)M',(n-1)N'}(k,0,0) \tag{11}$$

$$S'_{\sigma,m,n} = \frac{1}{k_1 k_2 N_1 N_2} \sum_{k=1}^{k_1 k_2} \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} F_{(m-1)M'+i,(n-1)N'+j}(k,i,j)^2 - S'^{\,2}_{\mu,m,n} \tag{12}$$

where, k_1 and k_2 are defined through $M' = k_1 N_1$ and $N' = k_2 N_2$ respectively. ($N_1 = N_2 = 8$ for MPEG-2).

These S-sequences can by used for higher level processing in image or video.

3. RESULTS

In this section, some results are presented with the proposed scheme. We used several test sequences and compare the estimated features against the actual values. Two QCIF test sequences: *akiyo* and *foreman* are considered here. First sequence doesn't contain any significant variations and second sequence exhibit large scene variations. The compressed version of these sequences is considered and both mean and standard deviation (variance) is extracted using the proposed scheme.

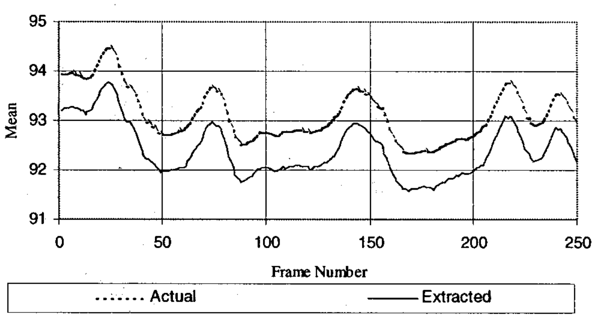

Figure 2: Comparison between the actual and the estimated mean (*akiyo*)

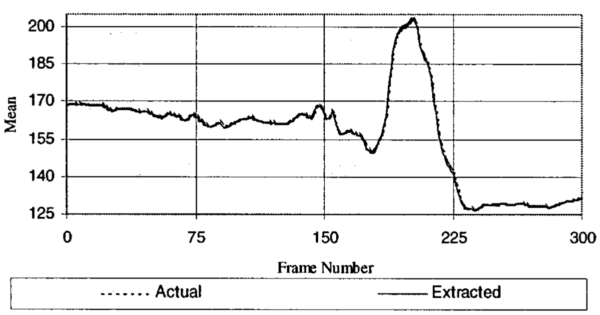

Figure 3: Comparison between the actual and the estimated mean (*foreman*)

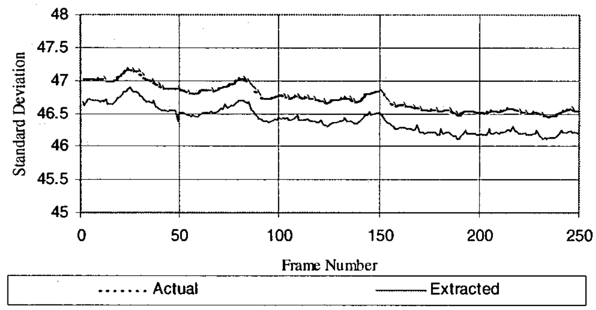

Figure 4: Comparison between the actual and the estimated standard deviation (*akiyo*)

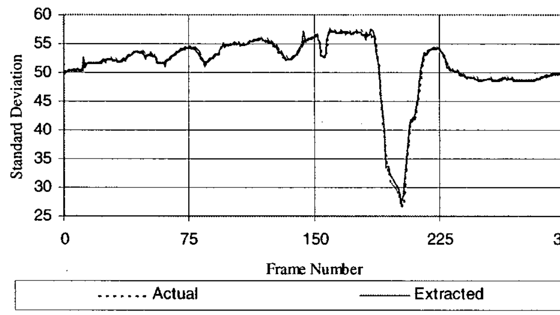

Figure 5: Comparison between the actual and the estimated standard deviation (*foreman*)

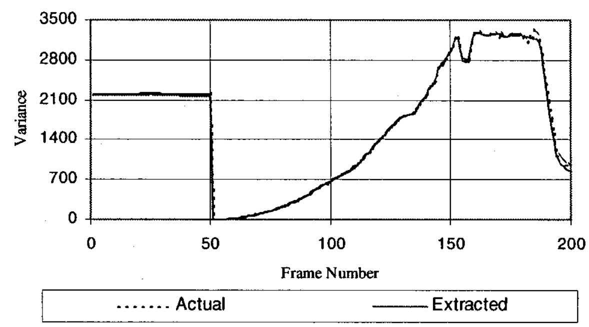

Figure 6: Comparison between the actual variance and the estimated variance for the fade-in sequence

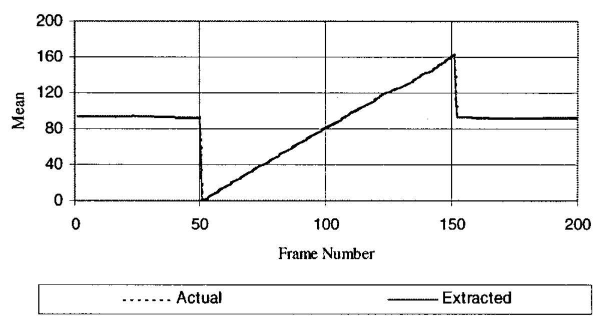

Figure 7: Comparison between the actual mean and the estimated mean for the fade-in sequence

Figure 2 and Figure 3 present the actual and extracted mean for *akiyo* and *forman* video sequences. Actual and extracted standard deviation are illustrated in Figure 4 and Figure 5 for the same sequences. These illustrations clearly show that extracted statistical features closely follows the actual values. Finally, this variance extraction technique is applied to identify fading and dissolving in MPEG-2 compressed domain with one of our algorithm proposed for uncompressed video [3]. In this algorithm [3], ratio between the second derivative of the variance curve and the first derivative of the mean curve is considered. The differentiation of this ratio is checked to identify fading and dissolving with a threshold T_{dis}. Figure 6 shows the behaviour of actual and extracted variance for a fade-in sequence. These clearly illustrate that extracted variance closely follows the actual variance and hence this can be used to identify fading/dissolving in compressed domain. Similar observations can also be seen in Figure 7 for mean.

Table 2 presents some of the results for fading/dissolving scene change detection with the proposed feature extraction technique along with the algorithms discussed in [3]. Results show that detected transition regions in compressed domain are equal or very close to the detected regions in uncompressed domain.

Actual transition region	Detected transition region (Uncompressed)	Detected transition region (Compressed)	Classification
446-496	446-497	446-497	dissolve
754-778	754-778	754-778	fade-out
804-868	804-868	804-868	dissolve
1010-1089	1010-1089	1012-1089	dissolve
1168-1232	1169-1232	1169-1233	dissolve
1356-1424	1356-1424	1356-1424	dissolve
1500-1550	1500-1550	1500-1552	fade-in
1620-1680	1620-1681	1620-1681	fade-in
1760-1840	1761-1840	1762-1840	fade-out
1920-1985	1920-1985	1920-1985	fade-out

Table 2: Dissolve/fade region identification

Furthermore, we used this variance extraction scheme to identify wiping in MPEG-2 compressed video and observed these scene changes were also identified accurately. Therefore, this feature extraction technique can be used effectively in compressed video to identify scene changes.

4. CONCLUSIONS

In this paper we presented a novel scheme to estimate variance of the images in MPEG-2 compressed video directly without full frame decompression. Results show that the estimated variance closely matches the actual variance. Furthermore, the proposed feature extraction technique was used to identify scene changes in compressed video without full frame decompression.

ACKNOWLEDGEMENTS

First author would like to express his gratitude and sincere appreciation to the university of Bristol and CVCP for providing financial support for this work.

REFERENCES

1. Wallace, G.K., 'The JPEG Still Picture Compression Standard', *Communications of the ACM*, Vol. 34, pp. 30-44, April 1991.
2. Sikora T., 'MPEG Digital Video-Coding Standards', *IEEE Signal Processing Magazine*, pp.82-100, September 1997.
3. Fernando, W.A.C., Canagarajah, C.N. and Bull, D. R., 'Fade and Dissolve Detection in Uncompressed and Compressed Video Sequences', *Proceedings of IEEE International Conference on Image Processing*, CD Paper Number 27AP2.10, Kobe, Japan, 1999.
4. Fernando, W.A.C., Canagarajah, C.N. and Bull, D. R., 'Wipe Scene Change Detection in Video Sequences', *Proceedings of IEEE International Conference on Image Processing*, CD Paper Number 27AP2.9, Kobe, Japan, 1999.
5. Alattar, A.M., 'Detecting Fade Regions in Uncompressed Video Sequences', *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, pp. 3025-3028, 1997.
6. Alattar, A. M., 'Wipe Scene Change Detector For Segmenting Uncompressed Video Sequences', *Proceedings of International Symposium on Circuits and Systems, Volume IV-Image and Video Processing*, pp. 249-252, 1998.
7. Yeo, B.L., 'Rapid Scene Analysis on Compressed Video', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 5, No 6, pp. 533-544, December-1995.
8. Chang, S.F., Messerschmitt, D.G., 'Manipulation and Composting of MC-DCT Compressed Video', *IEEE Journal of Selected Areas in Communications*, vol.13, pp.1-11, January 1995.